(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **06290210 A**

(43) Date of publication of application: **18 . 10 . 94**

(51) Int. Cl **G06F 15/38**

(21) Application number: **05075638**

(22) Date of filing: **01 . 04 . 93**

(71) Applicant: **SHARP CORP**

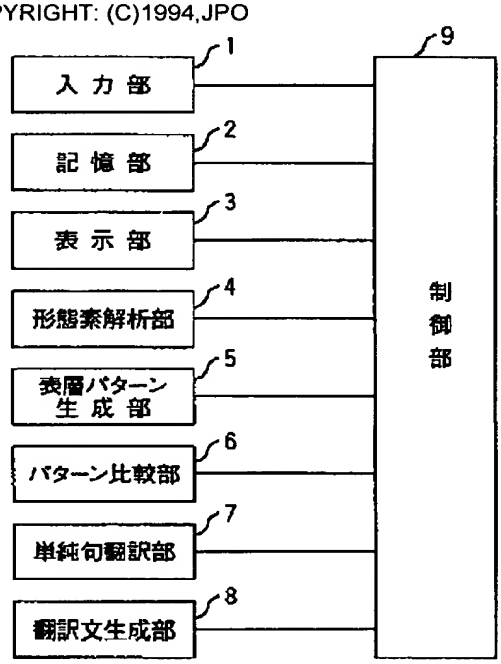
(72) Inventor: **MORISHITA TARO**
TSUBAKI KAZUHIRO
YAMAJI TAKAHIRO
KOBUCHI YASUJI

(54) TRANSLATING DEVICE FOR NATURAL LANGUAGE

(57) Abstract:

PURPOSE: To obtain a translation of high quality without analyzing the structure of a sentence and editing it.

CONSTITUTION: A storage part 2 stores a translation example data base. A morpheme analytic part 4 performs a morpheme analysis of an input sentence from an input part 1 to obtain a predicate. A surface layer pattern generation part 5 generates the surface layer pattern of the input sentence on the basis of the morpheme analytic result. A pattern comparison part 6 determines an index of the translation example data base on the basis of the predicate of the input sentence and further retrieves conversion patterns and translation examples under the index by using the surface layer pattern of the input sentence. A simple phrase translation part 7 translates a word string corresponding to category symbols of the surface layer pattern of the input sentence by referring to the translation examples and fill the empty field in the conversion pattern to obtain a character string pattern of a target language. A translation generation part 8 generates a complete translation on the basis of the character string pattern of the target language. Thus, the translation of high quality is obtained without analyzing the structure of the input sentence and editing it.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平6－290210

(43)公開日　平成６年(1994)10月18日

(51)Int.Cl.[5]	識別記号	庁内整理番号	Ｆ Ｉ	技術表示箇所
Ｇ０６Ｆ　15/38		Ｐ　7323－5Ｌ		
		Ｅ　7323－5Ｌ		

審査請求　未請求　請求項の数２　ＯＬ　（全 15 頁）

(21)出願番号　特願平5－75638

(22)出願日　平成５年(1993)４月１日

(71)出願人　000005049
シャープ株式会社
大阪府大阪市阿倍野区長池町22番22号

(72)発明者　森下　太朗
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内

(72)発明者　椿　和弘
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内

(72)発明者　山路　孝浩
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内

(74)代理人　弁理士　青山　葆　（外１名）

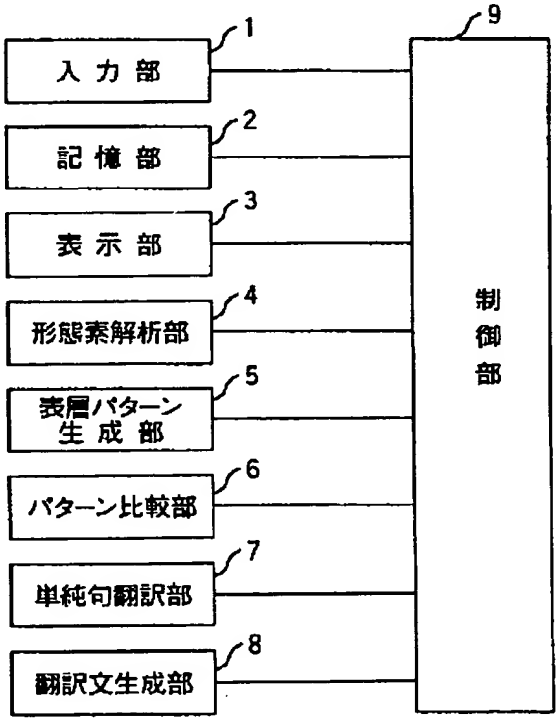
最終頁に続く

(54)【発明の名称】　自然言語の翻訳装置

(57)【要約】

【目的】　文の構造解析や編集をすることなく質の高い訳文を得る。

【構成】　記憶部２には対訳例文データベースを格納する。形態素解析部４は入力部１からの入力文を形態素解析して述語を得る。表層パターン生成部５は形態素解析結果に基づいて入力文の表層パターンを生成する。パターン比較部６は入力文の述語に基づいて対訳例文データベースのインデックスを決定し、更に入力文の表層パターンを用いて当該インデックス下に在る変換パターンおよび対訳例文を検索する。単純句翻訳部７は対訳例文を参照して入力文の表層パターンにおけるカテゴリ・シンボルに対応する単語列を翻訳して変換パターン内の空欄を埋めて目標言語の文字列パターンを得る。翻訳文生成部８は目標言語の文字列パターンに基づいて完全な翻訳文を生成する。こうして、入力文の構造解析や編集をすることなく質の高い訳文を得る。



【特許請求の範囲】
【請求項１】　入力部から入力された自然言語による文章に対して形態素解析部で形態素解析を行い、記憶部に格納されている訳例文データベースから入力文に対応する例文とその対訳との対である対訳例文を対訳例文検索部によって上記形態素解析結果に基づいて検索し、この検索された対訳例文に基づいて翻訳部で入力文章を目標言語に翻訳し、得られた翻訳結果を表示部に表示する自然言語の翻訳装置において、
上記形態素解析部による入力文に対する形態素解析結果に基づいて、上記入力文から、少なくとも用言および付属語の文字列とそれらに前後する単語列の構文カテゴリとによって文の表層的特徴を表した表層パターンを所定の手順で生成する表層パターン生成部を備えると共に、
上記対訳例文データベースに蓄積された各対訳例文には、この対訳例文における例文の上記表層パターンを予め付加し、
上記対訳例文検索部は、上記対訳例文データベースから該当する対訳例文を検索するに際して、上記表層パターン生成部によって生成された入力文の表層パターンと上記対訳例文に付加されている例文の表層パターンとの類似度を求めることによって、入力文に類似した例文を有する対訳例文を検索する構成に成したことを特徴とする自然言語の翻訳装置。
【請求項２】　請求項１に記載の自然言語の翻訳装置において、
上記記憶部に格納された対訳例文データベースには、用言の文字列パターンをルートノードとし、当該用言を用いた文から抽出された少なくとも当該用言および付属語の文字列パターンを当該ルートノードから分岐した各ノードとする木構造を成すと共に、上記各ノードの文字列パターンは親ノードの文字列パターンを詳細化した文字列パターンになっているインデックス木を設けて、このインデックス木におけるリーフノードの文字列パターンを上記対訳例文データベースのインデックスとし、
上記対訳例文検索部は、入力文から上記形態素解析部での形態素解析結果によって抽出された用言に基づいて、当該用言を表す文字列パターンのルートノードを有するインデックス木を検索し、この検索されたインデックス木を用いて上記対訳例文データベースのインデックスを得る構成に成したことを特徴とする自然言語の翻訳装置。
【発明の詳細な説明】
【０００１】
【産業上の利用分野】この発明は、自然言語で書かれた文章を自動的に翻訳する自然言語の翻訳装置に関する。
【０００２】
【従来の技術】従来より、機械翻訳装置としては、図７に示すような解析レベルに従って、辞書情報と多数の解析ルールを使用して目標言語との対応が取り易くなるレベルまで原言語による入力文(以下、原文と言う)の解析を行い、原文が表す意味的な内部構造を抽出するという解析プロセスを採用した所謂トランスファー方式によるものが主流である。
【０００３】すなわち、先ず、原文に対する形態素解析によって各単語に対する品詞列を求める。次に、構文解析によって上記品詞列に対する句構造を求める。そして、最後に単語や句の用法に関する種々のデータを使って意味解析を行って依存構造等の最終的な内部構造を得る。こうして、目標言語との対応が取り易くなるレベルまで解析されると、目標言語への変換規則を用いて同レベルの目標言語に変換し、そこから構文生成,形態素形成と生成プロセスを進めて目標言語を生成して行くのである。
【０００４】このように、従来の機械翻訳装置では、解析主導の翻訳プロセスが翻訳処理の前提となっている。ところが、上記従来の解析主導の翻訳システムには以下のような欠点がある。
【０００５】
（１）翻訳の専門家のような柔軟な意訳ができない
目標言語への変換規則は、通常機械的な置き換えによるものであり、分かりやすい表現の訳文にするための知識は反映されてはいない。したがって、得られる訳文は堅い表現になり、非常に分かりにくいものになっている。そのために、現行の機械翻訳装置では、マニュアルによって翻訳結果を“後編集”して分かりやすい訳文に修正したり、マニュアルによって入力文を“前編集”して機械翻訳装置が容易に処理可能な文型に書き換えたりしなければ、妥当な訳文を得ることができないのである。
【０００６】その結果、当然のことながら、人手を介することなく、翻訳の専門家が訳すようなレベルの“意訳”の訳文を得ることは極めて難しい。
【０００７】
（２）機械システムのメンテナンスや改良が困難
部分的に上記目標言語への変換規則等の解析ルールや経験則を増やして翻訳システムを改良しようとしても、全体の処理アルゴリズムに影響を及ぼしてしまうので変更に伴う負担が大きい。また、翻訳システムを修正できたとしてもヒューリスティックに依存する部分が多く、ヒューリスティックを統一的に制御する有効な手段を備えてはいないために、翻訳改善の対象となった文に対しては良好な改善結果が得られる一方で、別の文章に対しては翻訳精度が低下してしまうという事態が発生し易い。
【０００８】上述のように、解析主導の翻訳システムの欠点を解消すべく、近年、例文主導の翻訳システムが提唱されている。
【０００９】この例文主導の翻訳システムでは、入力文に最も類似した対訳例文(対訳を有する例文)を対訳例文データベースから検索し、この検索した対訳例文の対訳を利用して上記入力文に対する翻訳を得るようにしてい

る。この翻訳システムには、上記対訳例文データベースに対訳例文を追加するだけで性能向上を図れるという利点や、対訳例文によってカバーできる範囲内においては意訳レベルでの翻訳が実施できるという利点がある。

【００１０】

【発明が解決しようとする課題】しかしながら、上記従来の例文主導の翻訳システムには以下のような問題点がある。現在提唱されている例文主導の翻訳システムにおいては、上記対訳例文を検索する際に用いるキーとして文章の依存構造を予め用意しておくものが多い。このために、対訳例文の検索に際しては翻訳対象となる入力文の依存構造を求める必要がある。そして、そのためには、入力文章の形態素解析,構文解析,係り受け解析,意味解析を正確に行わなければならない。

【００１１】ところで、一般に、入力文に対する形態素解析および構文解析の際には多数の解析候補が得られる。そして、長く複雑な文章になるほど得られる解析候補の数が増大する。さらに、上記解析候補を絞り込むための意味解析においては、拠り所となる規則が存在しない。そこで、通常は多数の経験則を用意しておいて状況に応じて使い分けることになる。

【００１２】その結果、長く複雑な文章になるほど、上記得られた多数の解析候補から文意に沿った候補を一意に絞り込むことが困難になるのである。したがって、上記依存構造を有する対訳例文を上記対訳データベースに格納する上記例文主導の翻訳システムでは、係り受け関係の複雑な文章を正しく翻訳できる確率が低いという問題点がある。

【００１３】そこで、この発明の目的は、入力文の構文解析,係り受け解析および意味解析等の解析プロセスを適用する必要がなく且つ“後編集”および“前編集”を実施することなく、係り受けの複雑な入力文であっても質の高い訳文を得ることができる自然言語の翻訳装置を提供することにある。

【００１４】

【課題を解決するための手段】上記目的を達成するため、第１の発明の自然言語の翻訳装置は、入力部から入力された自然言語による文章に対して形態素解析部で形態素解析を行い、記憶部に格納されている訳例文データベースから入力文に対応する例文とその対訳との対である対訳例文を対訳例文検索部によって上記形態素解析結果に基づいて検索し、この検索された対訳例文に基づいて翻訳部で入力文章を目標言語に翻訳し、得られた翻訳結果を表示部に表示する自然言語の翻訳装置において、上記形態素解析部による入力文に対する形態素解析結果に基づいて、上記入力文から、少なくとも用言及び付属語の文字列とそれらに前後する単語列の構文カテゴリとによって文の表層的特徴を表した表層パターンを所定の手順で生成する表層パターン生成部を備えると共に、上記対訳例文データベースに蓄積された各対訳例文にはこの対訳例文における例文の上記表層パターンを予め付加し、上記対訳例文検索部は、上記対訳例文データベースから該当する対訳例文を検索するに際して、上記表層パターン生成部によって生成された入力文の表層パターンと上記対訳例文に付加されている例文の表層パターンとの類似度を求めることによって入力文に類似した例文を有する対訳例文を検索する構成に成したことを特徴としている。

【００１５】また、第２の発明は、第１の発明の自然言語の翻訳装置において、上記記憶部に格納された対訳例文データベースには、用言の文字列パターンをルートノードとし、当該用言を用いた文から抽出された少なくとも当該用言および付属語の文字列パターンを当該ルートノードから分岐した各ノードとする木構造を成すと共に、上記各ノードの文字列パターンは親ノードの文字列パターンを詳細化した文字列パターンになっているインデックス木を設けて、このインデックス木におけるリーフノードの文字列パターンを上記対訳例文データベースのインデックスとし、上記対訳例文検索部は、入力文から形態素解析部での形態素解析結果によって抽出された用言に基づいて当該用言を表す文字列パターンのルートノードを有するインデックス木を検索し、この検索されたインデックス木を用いて上記対訳例文データベースのインデックスを得る構成に成したことを特徴としている。

【００１６】

【作用】第１の発明では、入力部から入力された自然言語による文章に対して形態素解析部によって形態素解析が実施され、この形態素解析結果に基づいて、表層パターン生成部によって、入力文から、少なくとも用言および付属語の文字列とそれらに前後する単語列の構文カテゴリとによって文の表層的特徴を表した表層パターンが所定の手順で生成される。そうすると、対訳例文検索部によって、記憶部の対訳例文データベースに蓄積された各対訳例文に付加されている例文の表層パターンと上記表層パターン生成部で生成された入力文の表層パターンとの類似度が求められる。そして、この類似度に基づいて、入力文に類似した例文を有する対訳例文が検索される。

【００１７】以後、この検索された対訳例文に基づいて、翻訳部によって入力文章が目標言語に翻訳され、得られた翻訳結果が表示部に表示される。こうして、文全体の表層的特徴を表す表層パターンを用いた形態素レベルでの類似度算出のみによって、非常に簡単に入力文に対応する対訳例文を検索して質の良い翻訳が得られる。

【００１８】また、第２の発明では、入力部から入力された入力文が形態素解析部によって形態素解析され、この形態素解析結果に基づいて入力文の用言が抽出される。そうすると、対訳例文検索部によって、上記抽出された用言を表す文字列パターンのルートノードを有する

インデックス木が検索され、この検索されたインデックス木を用いて上記対訳例文データベースのインデックスが得られる。

【００１９】そして、こうして得られたインデックスを用いて、上記対訳例文検索部によって、入力文の表層パターンとの類似度算出の対象となる対訳例文候補が選出される。

【００２０】

【実施例】以下、この発明を図示の実施例により詳細に説明する。この発明における自然言語の翻訳装置は、文章の表層パターンを利用して対訳例文を検索する例文主導の翻訳システムを備えた翻訳装置である。

【００２１】図１は本実施例の自然言語の翻訳装置における概略ブロック図である。以下、便宜上、日本語による原文を英語に翻訳する場合を例に上記自然言語の翻訳装置を説明する。

【００２２】入力部１はキーボードや光学文字読み取り装置(ＯＣＲ)等の入力機器で構成されて、上記対訳例文や翻訳対象の文章等を入力する。記憶部２はＲＡＭ(ランダム・アクセス・メモリ)やＲＯＭ(リード・オンリ・メモリ)等のメモリおよびこのメモリを制御するメモリ制御手段で構成されて、単語辞書や対訳例文データベース等を格納する。表示部３はＣＲＴ(カソード・レイ・チューブ)等の表示機器で構成される。

【００２３】形態素解析部４は、記憶部２のメモリに格納されている単語辞書を引いて入力文章から単語列を切り出し、品詞列を生成する。さらに、テンスやアスペクト等の情報を得る。表層パターン生成部５は、形態素解析部４による形態素解析結果を用いて、入力部１からの入力文の表層パターンを生成する。パターン比較部６は、後に詳述するようにして、入力文章の表層パターンの候補と記憶部２の上記メモリに格納されている対訳例文データベースに用意されている表層パターンとの比較を行って、入力文章に最も類似した表層パターンを有する対訳例文を検索する。

【００２４】単純句翻訳部７は、複雑な埋め込み文のない名詞句(「本」,「その本」,「彼の本」,「美しい本」等)や、空列を含む助動詞列が後続する述語等の単純な語句を対象として、上記単語辞書のような簡単なルールのみに基づいて翻訳処理を実行する。この単純句翻訳部７は、上述した従来型の機械翻訳装置における一部の機能で代用可能であるために、ここでは詳細な説明は省略する。

【００２５】翻訳文生成部８は、上記目標言語における単語の並びやテンスおよびアスペクト等の情報から、目標言語による完全な翻訳文を生成する。尚、この翻訳文生成部８についても従来型の機械翻訳装置における一部の機能で代用可能であるために詳細な説明は省略する。制御部９は、上記入力部１,記憶部２,表示部３,形態素解析部４,表層パターン生成部５,パターン比較部６,単純句翻訳部７および翻訳文生成部８を制御して、入力文章の翻訳処理を実施する。

【００２６】すなわち、上記パターン比較部６で上記対訳例文検索部を構成し、単純句翻訳部７,翻訳文生成部８および制御部９で上記翻訳部を成すのである。

【００２７】本実施例の翻訳装置によって実施される翻訳処理の概略は、入力文の表層パターンを用いて入力文に最も類似した対訳例文を上記対訳例文データベースから検索し、検索された対訳例文における対訳を基にして入力文の翻訳文を得る処理である。以下、上記翻訳処理について順を追って詳細に説明する。尚、ここで言う表層パターンとは、文を特徴付ける単語とその他の部分単語列の構文カテゴリとによって表されるものである。

【００２８】先ず、上記記憶部２のメモリに格納される対訳例文データベースについて説明する。図２および図３は上記対訳例文データベースに関する説明図である。図２は上記対訳例文データベースのインデックス構造を示し、「ある」という動詞が述部となる和文を原文とする複数の対訳例文のインデックス構造を例示している。

【００２９】上記インデックスは、述語の終止形「ある」をルートノードとし、その述語「ある」を含む表層の文字列パターン「＊は＊が ある」,「＊には＊が ある」,「＊は＊に＊が ある」,…をルートノード以外のノードとする木構造で表現される。尚、上記表層の文字列パターンにおけるパターン要素は、各ノードに存在する述語「ある」に対する必須格,任意格,省略格の格助詞および特徴的な単語である。上記表層の文字列パターンは、リーフノードに行くほど詳細に記述され、子ノードの文字列パターン(例えば、「＊は＊と＊が ある」)は親ノードの文字列パターン(例えば、「＊は＊が ある」)を詳細化した文字列パターンになっている。

【００３０】そして、上述のようなルートノードを幹とする木構造を有するインデックス木の各リーフノードに係る上記表層の文字列パターンを上記対訳例文データベースのインデックスとし、このインデックスに対訳例文が対応付けられている。したがって、入力文から抽出した述語の終止形をルートノードとするインデックス木を上記入力文の表層の文字列に従って辿って行くことによって、該当する対訳例文を検索するためのインデックスを決定できるのである。

【００３１】図３は、「＊には＊が ある」という表層の文字列パターンを有する和文を原文とする対訳例文を蓄積した対訳例文データベースの構造例を示す。図３に示すように、上記対訳例文データベースは、上記インデックス(上記インデックス木のリーフノードに係る表層の文字列パターン),表層パターン,変換パターンおよび対訳例文からなる層構造を成している。

【００３２】ここで、上記表層パターンは本実施例の中心となるデータ構造であり、上述したように文を特徴付ける単語(以下、特徴単語と言う)とその他の部分単語列の構文カテゴリによって表される。ここで、上記特徴単

語とは、動詞,助詞および一部の特徴的な名詞であり、図２に示すインデックスにおける各ノードの文字列パターンに具体的に表記された単語に対応する。また、上記構文カテゴリとは、上記特徴単語に前後する単語列(すなわち、上記インデックスでは“＊”に対応する部分単語列)の簡単な句構造を表すものである。

【００３３】次に、上記表層パターンの構成法について説明する。

構文カテゴリ		カテゴリ・シンボル
単純名詞句	→	N
埋め込み文によって装飾された名詞句	→	VP・N
動詞句	→	VP
…		…

【００３４】上述のようにして構成される表層パターンを用いて、上記対訳例文データベースは次のように構成される。以下、図３に従って対訳例文データベースの構成について具体的に説明する。

【００３５】「＊には＊が ある」というインデックス下には、次のようなパターン１～パターン３と命名された３つの表層パターンが存在する。すなわち、

パターン１＝“N1にはN2が ある”

＝“単純名詞句１＋「には」＋単純名詞句２＋「が」＋「ある」”

パターン２＝“VP・N1にはN2が ある”

＝“連体修飾述句＋単純名詞句１＋「には」＋単純名詞句２＋「が」＋「ある」”

パターン３＝“VPにはNが ある”

＝“述句＋「には」＋単純名詞句＋「が」＋「ある」”

【００３６】さらに、各表層パターン下には、その表層パターンを有する和文を英訳する際に用いられる変換パターンが存在する。

例えば、　パターン１＝“N1にはN2が ある”

＝“単純名詞句１＋「には」＋単純名詞句２＋「が」＋「ある」”

に対しては、

変換パターン＝“There BE T(N2) in T(N1).”

が対応付けられており、

“「There」＋BE動詞＋単純名詞句２の翻訳結果＋「in」＋単純名詞句１の翻訳結果”

が変換されるべき英文のパターンであることを示している。

【００３７】ここで、上記変換パターンに見られる“T(x)”という表記は、句“x”に対応する単語列を上記単純句翻訳部７(図１参照)によって翻訳した結果を表す。例えば、CASE01に示す対訳例文の場合には、“x”は「庭」を表す単純名詞句であり、“T(x)”は「garden」である。また、“$Tc_h(x)$という表記は、CASE番号“h”を有する対訳例文の対訳英文を表す。例えば、CASE11に示す対訳例文の場合には、“x”は「彼が学会誌に発表した論文」を表す埋め込み文を含む名詞句であり、CASE番号

(１)　対象となる文の中心用言とそれに係る任意格を含めた格助詞,接続助詞および特徴的な名詞とを夫々抽出して上記特徴単語とする。

(２)　(１)で抽出された特徴単語に前後する部分単語列の上記構文カテゴリを設定する。そして、その設定された構文カテゴリを次のようにカテゴリ・シンボルに置き換える。

【００３８】つまり、上記変換パターンは一種のテンプレートとなっており、対応する表層パターンを構成する上記特徴単語に前後する部分単語列の翻訳結果で上記テンプレートの空欄を埋めることによって翻訳英文が得られるのである。

【００３９】上述のような構造を有する対訳例文データベースとして大量の対訳例文を蓄積しておけば、入力文章の表層パターンと類似若しくは一致した表層パターンを有する対訳例文を対訳例文データベースから検索することによって、質の高い翻訳文を得ることが容易に可能となるのである。

【００４０】ここで、上述のような表層パターンを用いて翻訳を実施することによって、次のような利点が得られるのである。

【００４１】(A)上記対訳例文データベースから入力文に類似若しくは一致する対訳例文を検索する際に実施される表層パターンのマッチングは、１次元的な形態素解析レベルでのパターンマッチングである。したがって、依存構造解析のように２次元的な解析を行う必要がない。具体的には、上記依存構造解析の場合には、入力文全体に対する係り受け解析および意味処理を含めた構文解析を必要とする。これに対して、表層パターンのマッチングの場合には、文字列のパターンマッチング,形態素解析および品詞列に対する極簡単なパターン認識処理しか必要とはしない。したがって、入力文章の解析処理が非常に単純なものとなる。

【００４２】このように、上記対訳例文の検索に伴う解析処理が簡単になることによって、従来型の例文主導の翻訳システムに比較して長く複雑な入力文章に対する翻訳処理時間が大幅に短縮される。

“１１”の対訳例文に記載された同じ和文に対する対訳英文を取り出すことによって、“$Tc_{11}(x)$”＝「the paper which he published in a scholar journal」が得られる。尚、上記CASExxは、具体的な例文と対訳との対から成る対訳例文を表す。例えば、CASE01の場合には、和文「庭には池がある」と対を成す英訳文は「There is apond in the garden」である。

【００４３】(Ｂ)上記従来型の例文主導による翻訳システムで実施される依存構造解析は、局所的に解析ルールを適用してマッチングを行い、得られた結果を積み上げるボトムアップ方式である。そのために、部分的には正しく構造が解析されているにも拘わらず、文章全体としては係り受け関係や句のまとまりが誤っている解析候補が生成される場合が多い。

【００４４】これに対して、上記表層パターンは文全体を規定したものであるために、表層パターンのマッチング処理に際しては巨視的に見た場合の翻訳の失敗を避けることができる。また、その結果、訳文候補の組み合わせの爆発を避けることができる。以上の理由から、本実施例における表層パターンを用いた翻訳システムによれば、長く複雑な文章に対する翻訳の精度が飛躍的に向上するのである。

【００４５】次に、上記入力部１から入力された入力文章から上記表層パターンを抽出し、記憶部２のメモリに格納された対訳例文データベースから上記入力文章に類似した対訳例文を上記抽出された入力文章の表層パターンに基づいて検索する対訳例文検索処理動作について説明する。

【００４６】図４および図５は、上記制御部９によって記憶部２,形態素解析部４,表層パターン生成部５およびパターン比較部６を制御して実施される対訳例文検索処理動作のフローチャートである。以下、図４に従って、上記対訳例文検索処理動作について詳細に説明する。

【００４７】ステップＳ1で、上記形態素解析部４によって、入力部１から入力された入力文“Ｓ”の形態素が解析されて単語列および品詞列が切り出され、テンスおよびアスペクト等の情報が得られる。そして、得られた入力文Ｓの単語列および品詞列から入力文Ｓの述語“Ｖ”が決定される。ステップＳ2で、上記パターン比較部６によって、上記ステップＳ1において決定された述語Ｖをキーワードとして、図２の構造を有して上記対訳例文に関連付けられた複数のインデックス木から当該述語Ｖと同じ文字列パターンをルートノード(以下、“ルートノードＶ”と言う)とするインデックス木が検索される。

【００４８】ステップＳ3で、さらに上記パターン比較部６によって、上記検索されたインデックス木におけるルートノードＶから分岐している各子ノードchild(V)の文字列パターンのパターン要素をキーワードとして、全子ノードchild(V)の文字列パターンと入力文Ｓの文字列とが比較される。ステップＳ4で、上記キーワードであるパターン要素が入力文Ｓの文字列中に在るような子ノードchild(V)が存在するか否かが判別される。その結果存在すればステップＳ5に進み、存在しなければ上記対訳例文データベース内に入力文Ｓに類似する対訳例文はないとして対訳例文検索処理動作を終了する。

【００４９】ステップＳ5で、当該子ノードを親ノード“Ｆ”とする。ステップＳ6で、上記パターン比較部６によって、子ノードchild(F)に係る上記パターン要素をキーワードとして、全子ノードchild(F)の文字列パターンと入力文Ｓの文字列とが比較される。ステップＳ7で、上記パターン要素が入力文Ｓの文字列中に在るような子ノードchild(F)が存在するか否かが判別される。その結果、存在すればステップＳ5に戻って当該子ノードchild(F)から分岐したノードに対する処理に移行する。一方、存在しなければステップＳ8に進む。

【００５０】ステップＳ8で、上記ノードＦはリーフノードであるから、このノードＦの文字列パターンが入力文Ｓに類似した対訳例文を検索する際のインデックスであると決定される。ここで、便宜上、上記インデックスを“＊P_1＊P_2＊…＊P_j＊…＊P_J･Ｖ”と表す。但し、“P_j(ｊ＝１～Ｊ)”はｊ番目のインデックス要素であり、“＊”は上記インデックス要素に前後する部分文字列である。ステップＳ9で、上記ステップＳ8において決定されたインデックスの文字列パターンにおけるインデックス要素が参照されて、入力文Ｓの文字列が上記インデックス要素と同じ文字の箇所で分割される。その際に、上記入力文Ｓの文字列に上記インデックス要素と同一の部分文字列が複数あるために分割箇所が一意に決まらない場合には、総ての分割候補が求められて保持される。ここで、上記分割候補がＩ個あるとした場合には、このＩ個の分割候補の集合$\{b_i\}$は次のように表される。

$\{b_i\}=\{conc(S_{ij}\cdot P_j)_{j=1\sim J}\}_{i=1\sim I}$

但し、S_{ij}：ｊ番目のインデックス要素P_jの直前に位置する“＊”に対応する部分文字列

【００５１】ステップＳ10で、分割候補番号ｉと表層パターン番号ｋとに“１”がセットされる。また、マッチング評価値Ｅkと最大マッチング評価値Ｅk'と最大マッチング評価値を呈する表層パターン番号k'と最大評価値を呈する分割候補番号i'に“０”がセットされる。ステップＳ11で、上記表層パターン生成部５によって、ｉ番目の分割候補b_iの各部分文字列$(S_{ij})_{j=1\sim J}$に対して形態素解析が実施されて、以下のような分割候補b_iの表層パターンbp_iが求められる。

$bp_i=[X_{ij}\cdot P_j]_{j=1\sim J}$

但し、X_{ij}：部分文字列S_{ij}を形態素解析して得られた品詞列$H_1,\cdots,H_r,\cdots,H_R$に対して割り当てられる上記カテゴリ･シンボル列

【００５２】上記カテゴリ･シンボル列X_{ij}の割り当ては、次のような割り当てルールを適用して実施される。

(ａ) 品詞H_Rが動詞,動詞に続く付属語,名詞に続く述語型助動詞である場合にはカテゴリ･シンボル“ＶＰ”を割り当てる。

(ｂ) 品詞H_Rが名詞,名詞に続く接辞であり、且つ、ｒ＜Ｒであるｒに対して連体形の動詞である品詞H_rが存在する場合には、カテゴリ･シンボル列“ＶＰ･Ｎ”を割

り当てる。
（ｃ） 品詞H_Rが名詞,名詞に続く接辞であり、且つ、ｒ＜Ｒであるｒに対して動詞である品詞H_rが存在しない場合にはカテゴリ・シンボル“N”を割り当てる。
【００５３】ステップＳ12で、上記対訳例文データベースから上記ステップＳ8において決定されたインデックス下に在るｋ番目の表層パターン(以下、任意のインデックス下に在る表層パターンをインデックス内表層パターンと言う)dp_kが読み出される。ここで、当該インデックス下にはＫ個のインデックス内表層パターンdp_kが在るものとすると、このＫ個のインデックス内表層パターンの集合$\{dp_k\}$は次のように表される。
$\{dp_k\}=\{[C_{kj}\cdot P_j]_{j=1\sim J}\}_{k=1\sim K}$
但し、C_{kj}：ｊ番目の上記特徴単語P_jの直前に位置するカテゴリ・シンボル列
つまり、上記インデックス内表層パターンは、入力文Ｓと同じ述語Ｖを含む入力文Ｓと同じ上記表層の文字列パターンを有する表層パターンであると言える。ステップＳ13で、上記パターン比較部６によって、上記ステップＳ11において求められた入力文Ｓの表層パターンbp_iのカテゴリ・シンボル列X_{ij}と上記ステップＳ12において読み出されたインデックス内表層パターンdp_kのカテゴリ・シンボル列C_{kj}とが、総てのｊについて比較される。その結果、$X_{ij}=C_{kj}$または$X_{ij}\fallingdotseq C_{kj}$であればステップＳ18に進む。一方、$X_{ij}\neq C_{kj}$であればステップＳ14に進む。ここで、上記“$X_{ij}\fallingdotseq C_{kj}$”とは、カテゴリ・シンボル$X_{ij}$あるいはカテゴリ・シンボル$C_{kj}$のうち何れか一方のヘッドフィーチャーが他方のカテゴリ・シンボルと一致する場合である。
【００５４】ステップＳ14で、上記インデックス内表層パターンdp_kの表層パターン番号ｋの内容が最大値“Ｋ”より小さいか否かが判別される。その結果最大値“Ｋ”より小さければステップＳ15に進み、そうでなければステップＳ16に進む。ステップＳ15で、表層パターン番号ｋの内容がインクリメントされてステップＳ12に戻り、次のインデックス内表層パターンの処理に移行する。ステップＳ16で、上記分割候補番号ｉの内容が最大値“Ｉ”より小さいか否かが判別される。その結果最大値“Ｉ”より小さければステップＳ17に進み、そうでなければステップＳ21に進む。ステップＳ17で、分割候補番号ｉの内容がインクリメントされてステップＳ１１に戻り、入力文Ｓの次の分割候補の表層パターンに対する処理に移行する。
【００５５】ステップＳ１８で、上記ステップＳ13での比較結果に基づいて、入力文Ｓ(分割候補b_i)の表層パターンbp_iとインデックス内表層パターンdp_kとの間のマッチング評価値Ｅｋが以下のようにして算出される。すなわち、先ず、上記分割候補b_iの表層パターンbp_iのカテゴリ・シンボル列X_{ij}と上記インデックス内表層パターンdp_kのカテゴリ・シンボル列C_{kj}との比較結果に基づいて、以下のようにマッチ度CE_{kj}が設定される。
【００５６】上記マッチ度CE_{kj}は次のように設定される。
(イ) カテゴリ・シンボル列X_{ij}とカテゴリ・シンボル列C_{kj}とが完全に一致する場合($X_{ij}=C_{kj}$)
例えば、X_{ij}及びC_{kj}が共に埋め込み文によって装飾された名詞句“ＶＰ・Ｎ”である場合には、マッチ度CE_{kj}に“１.０”を与える。
(ロ) カテゴリ・シンボル列X_{ij}とカテゴリ・シンボル列C_{kj}のヘッドフィーチャーとが一致する場合($X_{ij}\fallingdotseq C_{kj}$)
例えば、X_{ij}が単純名詞句“Ｎ”でC_{kj}が埋め込み文によって装飾された名詞句“ＶＰ・Ｎ”である場合には、マッチ度CE_{kj}に“０.５”を与える。
(ハ) カテゴリ・シンボル列X_{ij}のヘッドフィーチャーとカテゴリ・シンボル列C_{kj}とが一致する場合($X_{ij}\fallingdotseq C_{kj}$)
例えば、X_{ij}が埋め込み文によって装飾された名詞句“ＶＰ・Ｎ”でC_{kj}が単純名詞句“Ｎ”である場合には、マッチ度CE_{kj}に“０.５”を与える。
【００５７】こうして、総ての“ｊ”についてマッチ度CE_{kj}が与えられると(すなわち、分割候補b_iの表層パターンbp_iとインデックス内表層パターンdp_kとが一致あるいは類似すると)、ｊ個のマッチ度CE_{kj}の和が算出されて表層パターンbp_iとインデックス内表層パターンdp_kとの間のマッチング評価値Ｅｋが得られる。
【００５８】ステップＳ19で、上記記憶部２のメモリに現在保持されている最大マッチング評価値Ｅｋ’と上記算出されたマッチング評価値Ｅｋとが比較される。その結果、当該マッチング評価値Ｅｋの方が最大マッチング評価値Ｅｋ’よりも大きい場合にはステップＳ20に進む。一方、最大マッチング評価値Ｅｋ’以下であればステップＳ14に戻って、次のインデックス内表層パターンが在れば次のインデックス内表層パターンに対する処理に移行する。ステップＳ20で、上記記憶部２によって、メモリに格納されている上記最大マッチング評価値Ｅｋ’を呈する表層パターン番号ｋ’が当該表層パターン番号“ｋ”に更新され、最大マッチング評価値Ｅｋ’を呈する分割候補番号ｉ’が当該分割候補番号“ｉ”に更新され、そして最大マッチング評価値Ｅｋ’が当該マッチング評価値“Ｅｋ”に更新される。その後、ステップＳ14に戻って、次のインデックス内表層パターン在れば次のインデックス内表層パターンに対する処理に移行する。
【００５９】ステップＳ21で、入力文Ｓに係る総ての分割候補b_i(ｉ＝１～Ｎ)および上記対訳例文データベースにおける当該インデックス下に在る総てのインデックス内表層パターンdp_k(ｋ＝１～Ｋ)に関する検索処理が終了したので、最大マッチング評価値Ｅｋ’を呈するインデックス内表層パターン下に在る対訳例文が出力される。また、最大マッチング評価値Ｅｋ’を呈する分割候補

の表層パターンが出力される。こうして、入力文Ｓの表層パターンに類似したあるいは一致した表層パターンを有する対訳例文が出力されて、対訳例文検索処理動作を終了する。

【００６０】このようにして、入力文Ｓに類似あるいは一致した対訳例文が得られると、当該対訳例文と当該対訳例文上に在る上記変換パターンとを入力文Ｓに適用して目標言語の具体化された文字列パターンを得る。その際における入力文Ｓへの適用とは、当該変換パターン内における表記Ｔ(ｘ)に対応する当該分割候補b_i内における部分文字列S_{ij}の上記単純句翻訳部７による翻訳や；当該変換パターン内における表記“Tc_h(ｘ)”で指定された対訳例文を用いた部分翻訳を意味する。

【００６１】こうして、上記目標言語の具体化された文字パターンが得られると、上記翻訳文生成部８によって、形態素解析部４による形態素解析で得られたテンスおよびアスペクトに関する情報や訳文生成ルールに基づいて、目標言語に具体化された文字パターンの時制，人称および数等の表現の検査/修正が行われて完全な翻訳文が生成される。そして、生成された翻訳結果は表示部３に出力されて表示される。

【００６２】次に、本実施例における翻訳装置によって実施される例文主導の翻訳処理について、入力例文を上げて図１～図５を参照して順を追って具体的に説明する。

【００６３】和文による入力文Ｓ「彼が買った本には落丁があった」が入力部１から入力される。そうすると、形態素解析部４で形態素解析が行われて述語Ｖ「ある」が決定され、入力文Ｓの時制情報“過去”が得られる。…ステップＳ１

上記述語Ｖ「ある」がルートノードになっている図２に示すインデックス木が検索される。そして、この検索されたインデックス木の子ノードの文字列パターンと入力文Ｓ「彼が買った本には落丁があった」の文字列とが比較されて、インデックス「＊には＊が ある」が決定される。…ステップＳ２～ステップＳ８

【００６４】上記インデックス「＊には(P_1)＊が(P_2) ある」が参照されて、入力文Ｓ「彼が買った本には落丁があった」が分割される。その際に、インデックス要素Ｐ$_1$(＝「には」)とこれに続くP_2(＝「が」)とは入力文Ｓ中に各々一回しか出現しないので、分割候補はb_1唯一つだけ存在する。

b_1＝「彼が買った本(S_{11})/には(P_1)/落丁(S_{12})/が(P_2)/ある(Ｖ)」…ステップＳ９

次に、上記分割候補b_1内の部分文字列S_{11}(＝「彼が買った本」)及び部分文字列S_{12}(「落丁」)に対する形態素解析が実施される。そして、上記部分文字列S_{11}(＝「彼が買った本」)には上記割り当てルール(ｂ)が適用されてカテゴリ・シンボルX_{11}(＝ＶＰ・Ｎ)に変換される。一方、部分文字列S_{12}(＝「落丁」)には割り当てルール(ｃ)が適用されてカテゴリ・シンボルX_{12}(＝Ｎ)に変換される。その結果、上記入力文Ｓにおける分割候補b_1の表層パターンbp_1が次のように求められる。

bp_1＝“ＶＰ・ＮにはＮが ある”…ステップＳ10，ステップＳ11

【００６５】上記対訳例文データベースにおけるインデックス「＊には＊が ある」下にはパターン１，パターン２およびパターン３と命名された３つのインデックス内表層パターンdp_1，dp_2，dp_3が存在する。そこで、上記分割候補b_1の表層パターンbp_1と各インデックス内表層パターンdp_1，dp_2，dp_3の夫々とが比較される。

① bp_1とdp_1との比較

bp_1＝“ＶＰ・Ｎ(X_{11}) には Ｎ(X_{12}) が ある”

dp_1＝“ Ｎ1(C_{11}) には Ｎ2(C_{12}) が ある”

したがって、X_{11}≒C_{11} → マッチ度CE_{11}＝０．５

X_{12}＝C_{12} → マッチ度CE_{12}＝１．０

マッチング評価値Ｅ1＝１．５

② bp_1とdp_2との比較

bp_1＝“ ＶＰ・Ｎ(X_{11}) には Ｎ(X_{12}) が ある”

dp_2＝“ＶＰ・Ｎ1(C_{21}) には Ｎ2(C_{22}) が ある”

したがって、X_{11}＝C_{21} → マッチ度CE_{21}＝１．０

X_{12}＝C_{22} → マッチ度CE_{22}＝１．０

マッチング評価値Ｅ2＝２．０

③ bp_1とdp_3との比較

bp_1＝“ＶＰ・Ｎ(X_{11}) には Ｎ(X_{12}) が ある”

dp_3＝“ ＶＰ(C_{31}) には Ｎ(C_{32}) が ある”

したがって、X_{11}≠Ｃ３１

分割候補b_1の表層パターンbp_1と見出し内表層パターンdp_3とは別表層パターンである。…ステップＳ12～ステップＳ18

【００６６】①および②での比較結果により、

マッチング評価値Ｅ1(＝１．５)＜マッチング評価値Ｅ2(＝２．０)

であるから、インデックス内表層パターンdp_2が入力文Ｓに最も類似したインデックス内表層パターンであると確定される。その結果、類似対訳例文としてCASE11と命名された

例文「彼が学会誌に発表した論文には誤りがある」

対訳「There are some errors in the paper which he published in a scholar journal」

の対が出力される。さらに、上記入力文Ｓの表層パターン

bp_1＝“ＶＰ・ＮにはＮが ある”が出力される。

…ステップＳ19～ステップＳ21

【００６７】こうして、入力文Ｓ「彼が買った本には落丁があった」の類似対訳例文が得られると、この得られた対訳例文上に在る上記変換パターン

「There BE Ｔ(N2) in Tc_h(ＶＰ・N1).」

に入力文Ｓが次にように適用される。

Tc_h(ＶＰ・N1＝彼が買った本) → 「the book which

he bought」
T(N2＝落丁) → 「missing page」
但し、この場合には、変換パターン「There BE T(N2) in Tc_{12}(ＶＰ・N1).」下には、例えば、
CASE12 ＶＰ＝彼が買った
N1＝本 N2＝誤り
There are some errors in the book which he bought.
なる対訳例文が記述されているものとする。
【００６８】こうして、和文による上記入力文S「彼が買った本には落丁があった」の目標言語(英語)に具体化された次のような文字列パターン記述が得られる。
「There BE missing page in the book which he bought.」
以後、この目標言語に具体化された文字列パターンと上記時制情報とに基づいて、上記訳文生成ルールを適用して、目標言語による翻訳文
「There were some missing pages in the book which he bought.」
を得るのである。
【００６９】上述の例では、説明の便宜を図るためにごく簡単な係り受け構造しか持たないような入力文Ｓの翻訳プロセスについて述べているが、更に複雑な係り受け構造を有する文章に対しても適切な翻訳文を得ることが可能である。例えば、以下のような入力文
「ハードウェアの構成は、本体とKBD,FDが一体になっているスタンドアロン型と、本体と一部が分離しているデスクトップ型の２種類があります。」
は、並列句が多く係り受け関係が複雑である。したがって、入力文章を一から解析する従来の解析主導の翻訳システムや依存構造を用いた例文主導の翻訳システムでは、入力文の解析段階で正しい解析結果を得ることが極めて困難である。したがって、高い翻訳精度は得られず、翻訳の専門家のような意訳ができず翻訳の質は低い。
【００７０】ところが、本実施例によれば、以下のように高精度で且つ質の高い翻訳文が得られるのである。すなわち、図６に示すように、上記対訳例文データベースに、
・インデックス
“＊は＊と＊の２＊が ある”
・インデックス内表層パターン
“N1はＶP1・N2とＶP2・N3の２N4がある”
・変換パターン
「There are two N4 of N1 : N2 and N3.
In the N2, Tc(ＶP1).
In the N3, Tc(ＶP2).」
・対訳例文
N1＝推論の方式 N2＝帰納法 N3＝演繹法 N4＝種類
ＶP1＝事実から規則を導く ＶP2＝規則から事実を導く
「There are two kinds of inference method : induction and
deduction.
In the induction, rules are infered from facts.
In the deduction, facts are infered from rules.」
を格納しておく。
【００７１】上記入力部１から上記入力文S「ハードウェアの構成は、本体とKBD,FDが一体になっているスタンドアロン型と、本体と一部が分離しているデスクトップ型の２種類があります。」が入力されると、上述のように形態素解析部２によって述語V「ある」が決定される。そして、上記パータン比較部５によってルートノードＶ“ある”のインデックス木が検索され、入力文Ｓの文字列に対応するインデックス“＊は＊と＊の２＊が ある”が求められる。こうして、上記対訳例文データベースのインデックスが決定されると、上述と同様に、決定されたインデックス下に在るインデックス内表層パターン，変換パターンおよび対訳例文を用いて入力文Ｓの目標言語に具体化された文字列パターン記述が得られるのである。
【００７２】このように、長く複雑な係り受けを有する入力文章であっても、その入力文章の表層パターンと同じ表層パターンを呈する対訳例文を対訳例文データベースに登録しておくだけで、翻訳生成に失敗することはないのである。また、長い文章の場合には、文意を取り易いように変換パターンおよび対訳例文の対訳を夫々複数に分割して(図６の場合には３つに分割)意訳するパターンで記述しておくことによって、専門家による翻訳に近い意訳が可能となる。
【００７３】上述のように、本実施例では、入力文章の表層の文字列パターンのマッチングおよび入力文章の文字列における上記特徴単語に前後する部分単語列の上記構文カテゴリのマッチングのみを実施すればよく、入力文章を解析して得られた複雑な依存構造によるマッチングを実施する必要がない。したがって、任意格や並列句を含む複雑な係り受け構造を有する入力文章にも容易に対処できる。
【００７４】尚、本実施例の翻訳装置では上記対訳例文をどれだけ網羅するかによって翻訳性能が決まる。一方、文の表層の文字列パターンの木を使用して対訳例文データベースをインデキシングするようにしている。したがって、本実施例の翻訳装置によれば、文法の専門家でなくとも系統的に対訳例文を増やして行くことが可能であり、翻訳性能の向上や翻訳システムの改良やメンテナンスを容易に実施できる。
【００７５】この発明における対訳例文検索処理動作のアルゴリズムは図４および図５に示すフローチャートに限定されるものではない。また、上記対訳例文データベースの具体的構成は、図３および図６に示すような構成

に限定されるものではない。

【００７６】

【発明の効果】以上より明らかなように、第１の発明の自然言語の翻訳装置は、形態素解析部による入力文の形態素解析結果に基づいて、上記入力文から、少なくとも用言および付属語の文字列とそれらに前後する単語列の構文カテゴリとによって文の表層的特徴を表す表層パターンを表層パターン生成部によって生成し、対訳例文検索部によって、上記表層パターン生成部で生成された入力文の表層パターンと上記対訳例文に付加された例文の表層パターンとの類似度を求めることによって入力文に類似した例文を有する対訳例文を検索するようにしたので、形態素レベルでの類似度算出によって上記対訳例文データベースから容易に該当する対訳例文を検索し、この検索された対訳例文を用いて例文主導の翻訳処理を実施できる。

【００７７】したがって、この発明によれば、入力文の構文解析,係り受け解析および意味解析等の２次元的な解析プロセスを適用することなく、“後編集”および“前編集”の実施の必要のない例文主導の翻訳処理を非常に簡単に且つ短時間に実施できるのである。

【００７８】さらに、その際における上記対訳例文検索部による類似度算出は、文全体の表層的特徴を表した表層パターンを用いて実施される。したがって、この発明によれば、係り受けの複雑な入力文であっても質の高い訳文を容易に得ることができる。

【００７９】また、第２の発明の自然言語の翻訳装置は、記憶部に格納された対訳例文データベースに、用言の文字列パターンをルートノードとし、当該用言を用いた文から抽出された少なくとも当該用言および付属語の文字列パターンを各ノードとする木構造を有するインデックス木を設けて、このインデックス木のリーフノードの文字列パターンを上記対訳例文データベースのインデックスとし、上記対訳例文検索部は、上記形態素解析部での形態素解析結果によって抽出された用言に基づいてインデックス木を用いて上記対訳例文データベースのインデックスを得るような構成にしたので、得られたインデックスに基づいて、上記対訳例文検索部による類似度計算の対象となる対訳例文候補を容易に選出できる。

【００８０】したがって、上記対訳例文検索部は、上記インデックスに基づいて選出された対訳例文候補に付加されている上記表層パターンに付いてのみ上記入力文の表層パターンとの類似度を求めればよく、入力文に類似した例文を有する対訳例文の検索を更に容易に且つ短時間に実施できる。

【図面の簡単な説明】

【図１】この発明の自然言語の翻訳装置におけるブロック図である。

【図２】図１における記憶部に格納された対訳例文データベースを検索する際に使用されるインデックス木の説明図である。

【図３】対訳例文データベースの構成例を示す図である。

【図４】対訳例文検索処理動作のフローチャートである。

【図５】図４に続く対訳例文検索処理動作のフローチャートである。

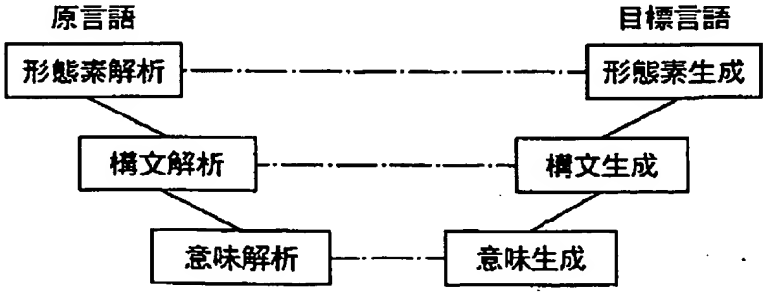
【図６】図３とは異なる対訳例文データベースの構成例を示す図である。

【図７】解析主導の翻訳プロセスにおける解析レベルの説明図である。

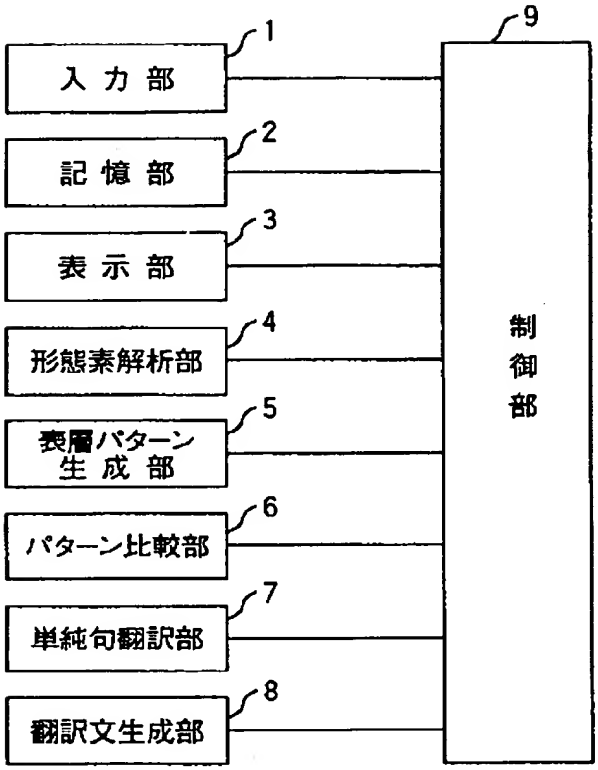
【符号の説明】

１…入力部、２…記憶部、３…表示部、４…形態素解析部、５…表層パターン生成部、６…パターン比較部、７…単純句翻訳部、８…翻訳文生成部、９…制御部。

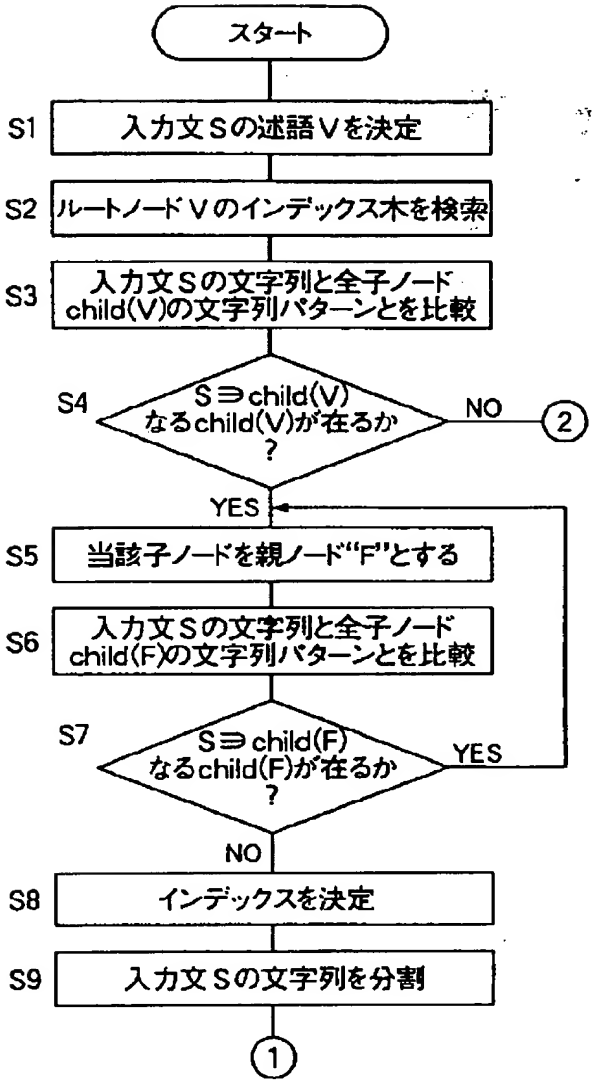
【図７】



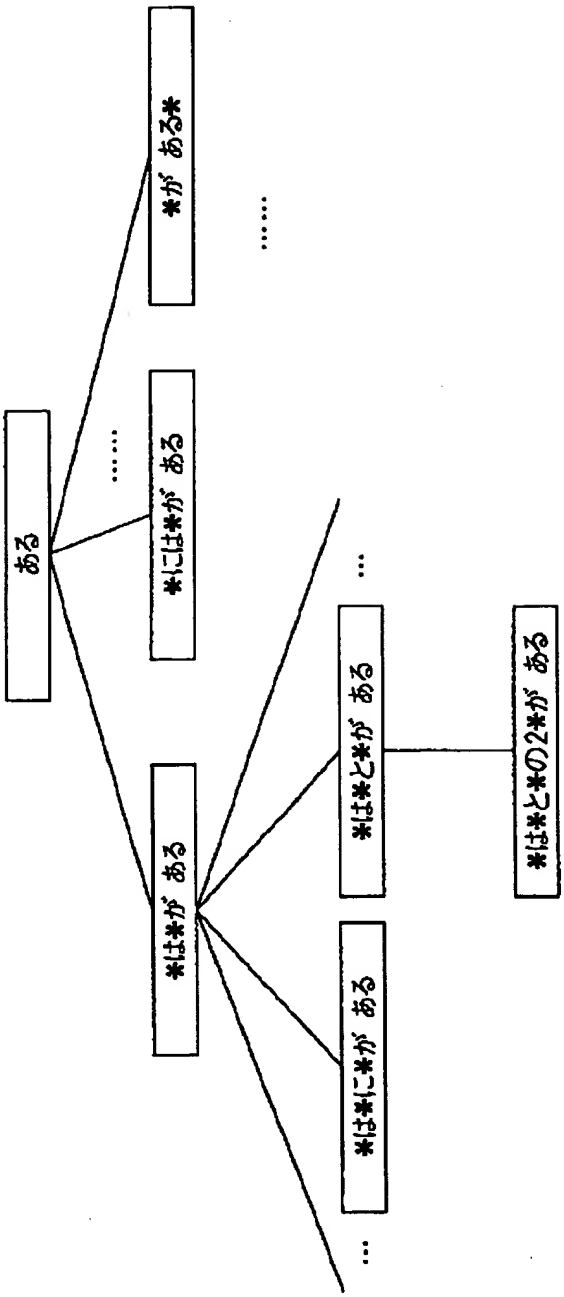
【図1】


1
入力部
2
記憶部
3
表示部
4
形態素解析部
5
表層パターン生成部
6
パターン比較部
7
単純句翻訳部
8
翻訳文生成部
9
制御部

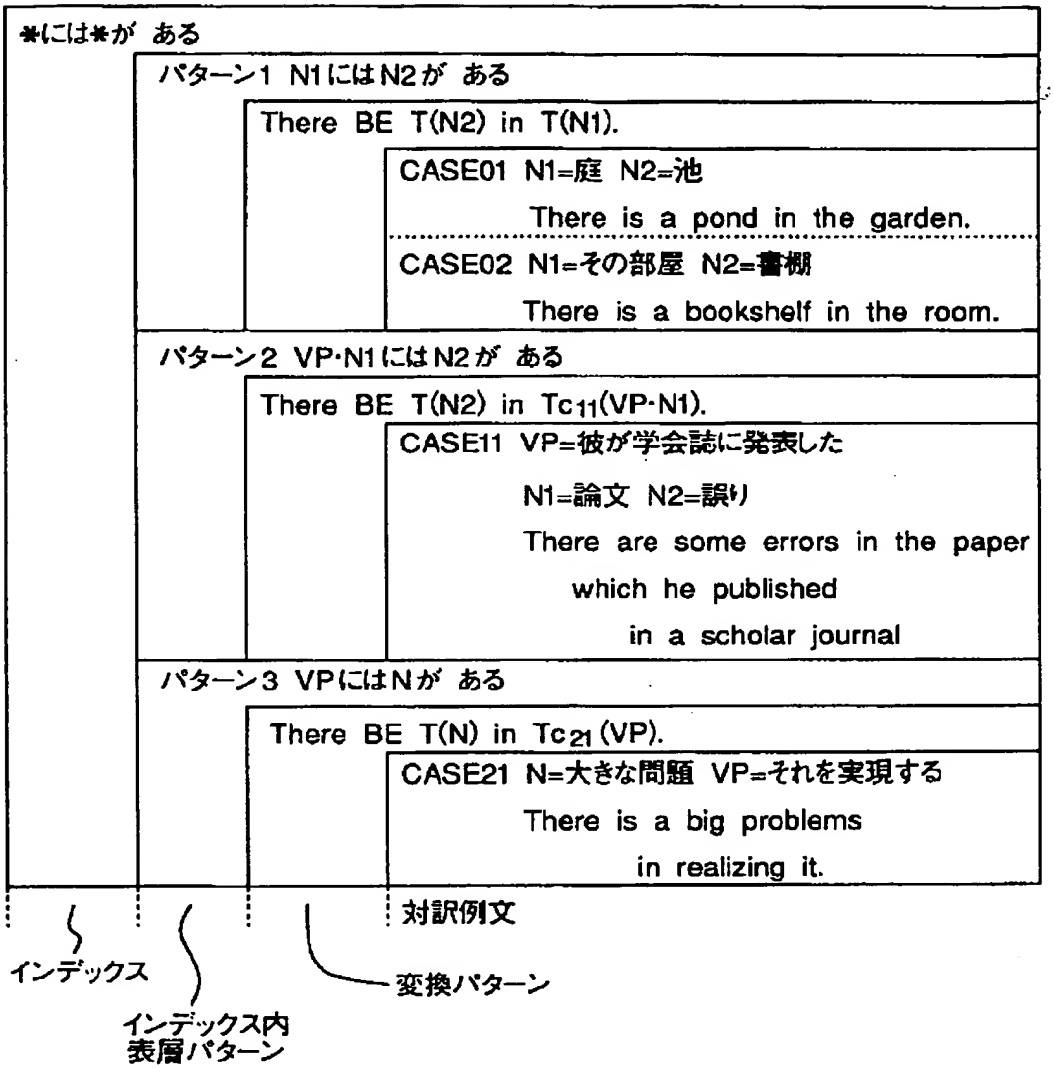
【図4】


スタート
S1
入力文Sの述語Vを決定
S2
ルートノードVのインデックス木を検索
S3
入力文Sの文字列と全子ノードchild(V)の文字列パターンとを比較
S4
S⊇child(V)なるchild(V)が在るか？
NO
2
YES
S5
当該子ノードを親ノード"F"とする
S6
入力文Sの文字列と全子ノードchild(F)の文字列パターンとを比較
S7
S⊇child(F)なるchild(F)が在るか？
YES
NO
S8
インデックスを決定
S9
入力文Sの文字列を分割
1

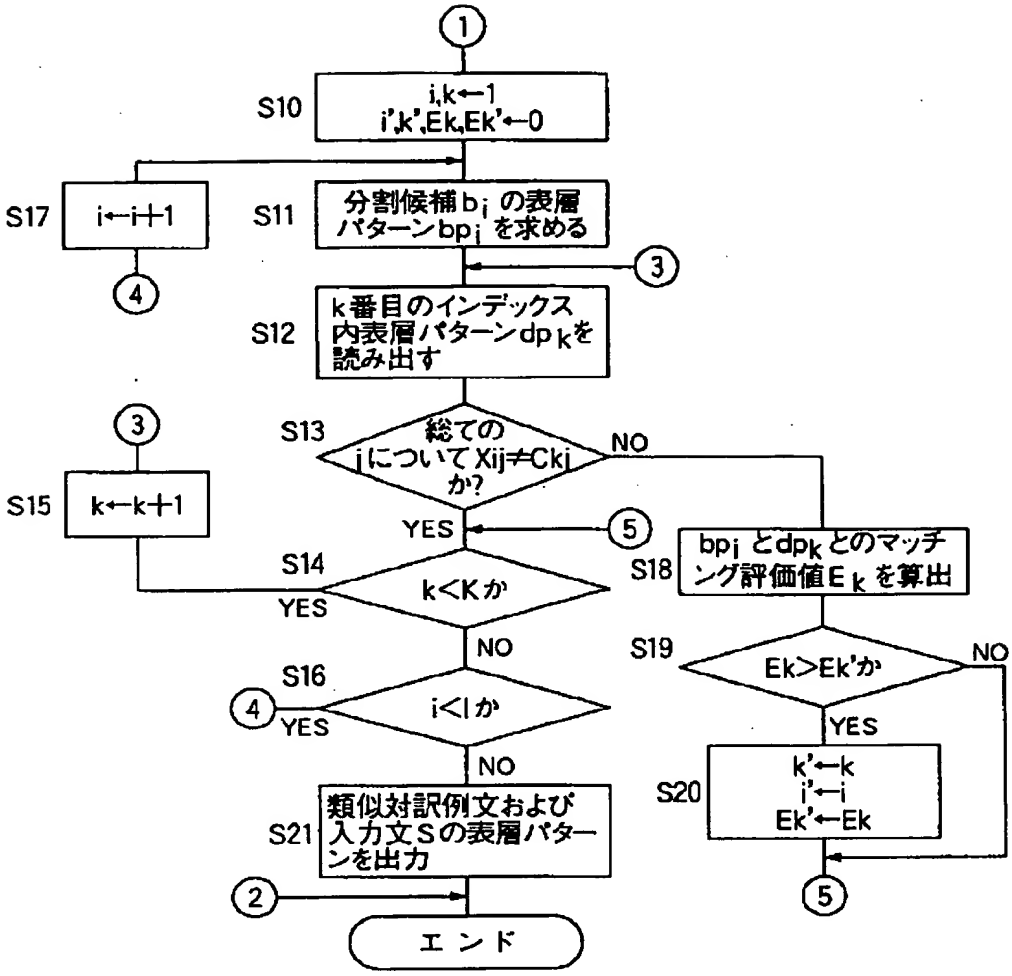
【図2】



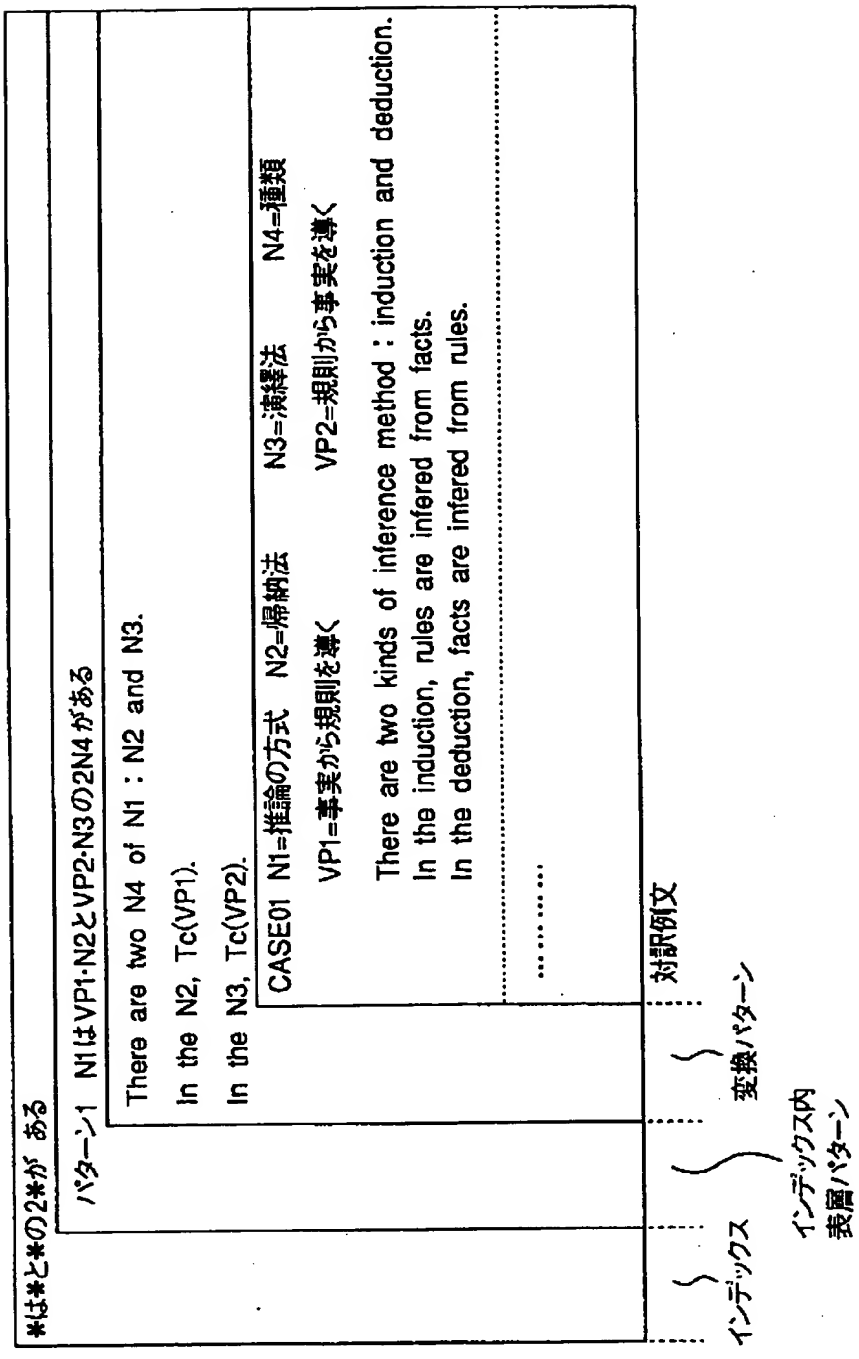
【図3】


*には*が ある
パターン1 N1にはN2が ある
There BE T(N2) in T(N1).
CASE01 N1=庭 N2=池
There is a pond in the garden.
CASE02 N1=その部屋 N2=書棚
There is a bookshelf in the room.
パターン2 VP·N1にはN2が ある
There BE T(N2) in Tc11(VP·N1).
CASE11 VP=彼が学会誌に発表した
N1=論文 N2=誤り
There are some errors in the paper
which he published
in a scholar journal
パターン3 VPにはNが ある
There BE T(N) in Tc21(VP).
CASE21 N=大きな問題 VP=それを実現する
There is a big problems
in realizing it.
対訳例文
インデックス
インデックス内
表層パターン
変換パターン

【図5】


①
S10
i,k←1
i',k',Ek,Ek'←0
S17
i←i+1
S11
分割候補 b_i の表層パターン bp_i を求める
③
④
S12
k番目のインデックス内表層パターン dp_k を読み出す
③
S13
総てのjについて $X_{ij} \neq C_{kj}$ か?
NO
S15
k←k+1
YES
⑤
S14
k<Kか
YES
S18
bp_i と dp_k とのマッチング評価値 E_k を算出
NO
S19
Ek>Ek'か
NO
S16
④
i<Iか
YES
YES
k'←k
i'←i
Ek'←Ek
S20
NO
S21
類似対訳例文および入力文Sの表層パターンを出力
⑤
②
エンド

【図6】



フロントページの続き

(72)発明者　小渕　保司
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内